

特許文書を対象とした因果関係抽出に基づく発明の新規用途探索 Search for New Application of Invention based on Extraction of Causality from Patents

太田 貴久^{*1}
Takahisa Ota

南 拓也^{*1}
Takuya Minami

山崎 祐介^{*1}
Yusuke Yamazaki

奥野 好成^{*1}
Yoshishige Okuno

田辺 千夏^{*1}
Chinatsu Tanabe

酒井 浩之^{*2}
Hiroyuki Sakai

坂地 泰紀^{*3}
Hiroki Sakaji

^{*1} 昭和電工株式会社
Showa Denko K.K.

^{*2} 成蹊大学
Seikei Univ.

^{*3} 東京大学
The Univ. of Tokyo

In this research, we propose a method for searching for new application of invention. The proposed method consists of two steps. The first step is to extract technical features and their effects from patents. The second step is to search for patents with similar technical features and dissimilar effect to a specific query. This method was demonstrated to be useful in searching for new application of invention.

1. はじめに

特許は、発明の公開を代償に、発明者へ独占排他権を付与する産業上重要な権利である。その際に公開される特許文書には、権利の範囲だけでなく、具体的な実現方法等の様々な技術的な情報が含まれている。特許法の目的でもある産業の発達を目指すためには、これらの情報を積極的に活用する必要がある。

一方、近年、知財を積極的に活用する取り組みの 1 つに「IP ランドスケープ」と呼ばれる概念がある[特許庁 2017]。IP ランドスケープとは、従来の事業保護を目的とした特許分析とは異なり、知財による事業拡大を目的とした分析、すなわち、将来を見据えた事業に資する知財分析を意味する。

このような背景のもと、本研究では、最も典型的な知財であり、かつ大量の情報を含む特許を用いて、事業拡大を目的とした分析を支援するための手法を提案する。具体的には、既存技術の新たな展開先(すなわち、既存発明の新しい用途)を探索する手法を提案する。

2. 従来研究

従来、特許文書から手段(原因)と効果(結果)の対を含む因果関係文を、コンピュータを用いて抽出する技術が知られている[日本特許 2011]。しかしながら、本特許[日本特許 2011] は、特許文書から手段と効果を抽出することにとどまっており、抽出した情報(因果関係文)の活用方法については触れられていない。

また、前記特許の技術を利用した研究として Nonaka らが提案する特許マップ自動生成システムがある[Nonaka 2010]。本研究は、特許文書から抽出した手段と効果を用いて、技術動向を把握するための図(特許マップ)を自動的に生成する研究である。しかしながら、本研究によって得られる特許マップは、指定した技術分野(特許集合)の範囲で生成されるものである。このような特許マップは、「ある分野における既知の効果に対して、未知の手段を把握する」ことには役立つが、技術分

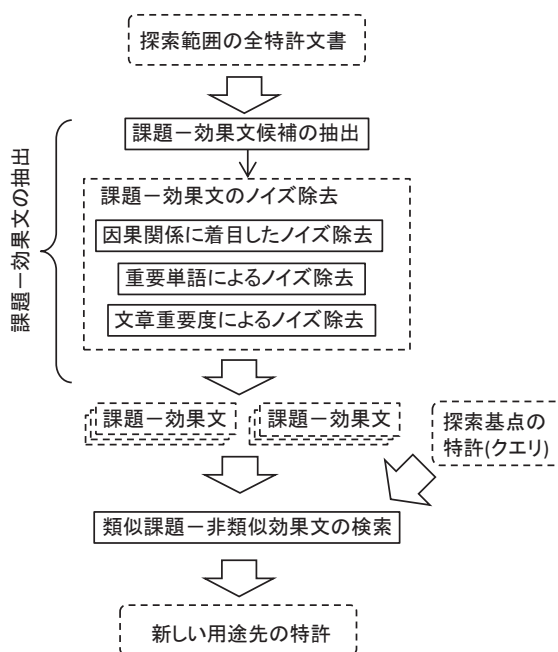


図 1: 提案手法の概要

野の枠を超えた、事業拡大を目指した分析を行えるものではない。

また、特許文書から手段と効果の因果関係を抽出し、手段が類似の因果関係は同様の効果を有するであろうという推測をコンピュータで行う研究が知られている[石川 2007]。しかしながら、石川らの研究[石川 2007]は、過去の特許文書に記載されている手段と効果の因果関係の範囲で、技術者の発想支援を行うものであり、事業に資する提案するようなものではない。

これらの研究に対して、本研究では、既存の技術が利用できる新たな技術分野(用途先)の発見を支援する。これにより、技術分野の枠を超えた事業に資する知財分析を提供することができる。

3. 提案手法

図 1 に本研究で提案する用途探索手法の概要を示す。

連絡先: 太田貴久, 昭和電工株式会社 法務・知的財産部, 東京都港区芝大門 1-13-9, (TEL) 03-5470-3256, (FAX) 03-3433-0960, ota.takahisa.xiuec@showadenko.com

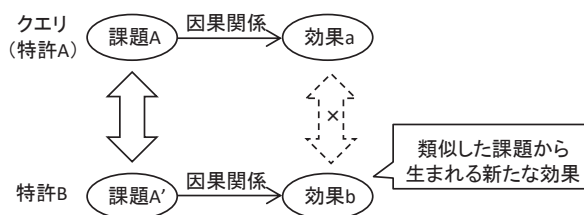


図 2: 類似課題－非類似効果

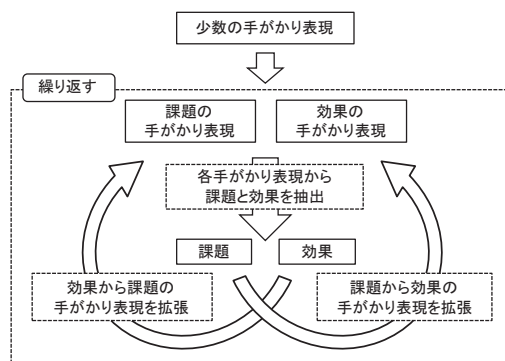


図 3: 坂地らの手法 [坂地 2010] の概要

図 1 のように、提案手法は、大きく分けると、「課題－効果文の抽出」と「類似課題－非類似効果文探索」の 2 つのステップに分けられる。

提案手法の最初のステップは、探索範囲のすべての特許文書(例えば、特定年代のすべての公開特許公報)から、課題－効果文の抽出を行う。ここで、課題－効果文とは、発明における最終的な因果関係を表す文であり、発明が達成する効果(以下、単に効果)と、その効果を達成するために解決する必要のあった技術課題(以下、単に課題)の組で表される情報である(従来研究では、課題を「手段」と呼んでいる)。例えば、「〇〇層を形成するため、熱伝導率を低減できる」という文が課題－効果文である。ここで、「〇〇層を形成するため」が課題を表し、「熱伝導率を低減できる」が効果を表す。効果の原因となる課題を抽出することで、単なる発明の構成要素ではなく、その発明の技術的な特徴を捉えることが可能となる。

課題－効果文の抽出ステップでは、課題－効果文の候補を抽出した後、3 種類のノイズ除去手法を適用することで、精度(適合率)の高い課題－効果文を抽出する。本研究は、新規用途探索を行う際に、分析者に新たな気付きを与えることを目的としており、技術を俯瞰することが目的ではない。そのため、すべての特許から漏れなく課題－効果文を抽出できなくても問題ない。すなわち、本研究の課題－効果文抽出においては、再現率より適合率(精度)の方が重要である。そこで、提案手法では、再現率を犠牲にしてもノイズとなるような課題－効果文を除去するようにしている。

提案手法の 2 つ目のステップは、類似課題－非類似効果文探索ステップである。本ステップは、探索の基点とする既存の課題－効果(以下、クエリと呼ぶ)をもとに、探索範囲の特許集合から新しい用途先(の候補)となる特許を探索するステップである。なお、クエリとする課題－効果は、課題と効果を表すようなテキストであればどのような指定を行ってもよい(必ずしも特許の文章を用いなくとも良い)。

ここで、探索するべき用途先の特許について考える。例えば、自社特許の新しい用途先を探索する場合、低コストで参入可能な用途先とは、自社の技術的特徴をそのまま応用できる技術分

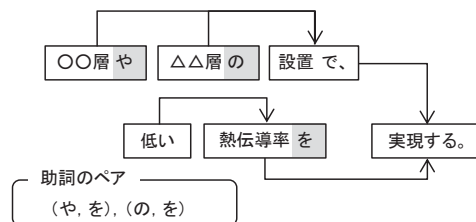


図 4: 坂地らの手法 [坂地 2011] の概要

野である。すなわち、前ステップで抽出した課題－効果文において、課題が類似している技術分野を探索できることが好ましい。一方、誰もが容易に想到する分野は好ましくない(特許文書の場合、公開までにタイムラグがあるため、すでに他社に取り組みられている可能性がある)。すなわち、前ステップで抽出した課題－効果文において、類似しない(意外性の高い)効果を持つ技術分野が好ましい(図 2)。

以上の理由により、提案手法の 2 つ目のステップでは、クエリと「類似した課題」を持ち、かつ「類似していない効果」を持つ特許を、探索範囲の特許集合から検索する。

提案手法は、以上の 2 ステップによって、類似している課題を持つ特許において、課題と対応している効果が類似していない特許を抽出する。これにより、クエリとした技術の新しい用途先を探索することが可能となる。以下では、各ステップの詳細について説明する。

3.1 課題－効果文候補の抽出

課題－効果文の抽出ステップの最初の処理では、探索範囲のすべての特許文書から、課題－効果文の候補を抽出する。本研究では、課題－効果文候補の抽出手法として、坂地らの手法(Cross-Bootstrapping) [坂地 2010] を用いた(以下、本手法を、坂地の手法(2010)と呼ぶ)。

坂地らの手法(2010)では、特許明細書の【発明の効果】は、課題と効果が同じ文に出現することが多いことを利用した手法である。坂地らの手法(2010)では、「手がかり表現」と呼ぶ、課題および効果を現す表現の周辺に出現する特徴的な表現に着目する。例えば、「本発明は、無効スペースとなる冷凍サイクル関連機器の配置構成を工夫することで、使用者が使いやすい位置の収納容積や収納性を高めることができる。」という文では、前の下線部が課題を、後の下線部が効果を現している。本例では、課題の直後の「ことで、」と、効果の直後の「ことができる。」が手がかり表現にあたる。坂地らの手法(2010)では、このような手がかり表現から課題－効果文を抽出する。簡単に手法を説明すると、坂地らの手法(2010)では、「手がかり表現から課題－効果文を抽出」するステップと「課題－効果から手がかり表現を拡張」するステップを交互に繰り返すブートストラップ手法である(図 3)。

3.2 因果関係に着目したノイズ除去

前の処理で抽出した課題－効果文の候補には多くのノイズが含まれる。例えば、坂地らの手法(2010)では、手がかり表現として「(課題)であり、(効果)である。」というパターンを用いて課題－効果文が抽出されることがある(この手がかり表現は、一見、完全な誤りに見えるが、「本発明の特徴は、〇〇層を設けたことであり、これにより、熱伝導率を低減できた点である。」という文も存在するため、完全な誤りとは言えない)。提案手法では、このようなノイズとなる課題－効果文を 3 つの方法で除去する。

1 つ目のノイズ除去の方法は、助詞のペアを用いたノイズ除去である。具体的には、坂地らの手法[坂地 2011]を用いる(以下、本手法を、坂地の手法(2011)と呼ぶ)。

坂地らの手法(2011)では、課題末尾の文節と効果末尾の文節のそれぞれに直接係る文節の助詞を抽出し、そのペアを基に因果関係を含む文か否かを判定する手法である。例えば、図4のように「〇〇層や△△層の設置で、低い熱伝導率を実現する。」という文の場合、課題末尾の文節である「設置で」に係る「〇〇層や」と「△△層の」と、効果末尾の文節である「実現する。」に係る「熱伝導率を」において、(課題側の文節の助詞、効果側の文節の助詞)のすべてのペアを用いて因果関係を判定する。この場合、(や、を)と(の、を)が因果関係の判定に用いられる。坂地らの手法(2011)では、事前に因果関係を含むと人手で判定した文の助詞のペアを教師データとして、SVMを用いて因果関係の判定を行っている。

3.3 重要単語に基づくノイズ除去

2 つ目のノイズ除去の方法は、重要単語に基づくノイズ除去である。最初のノイズ除去では、因果関係の有無に着目してノイズ除去を行った。しかしながら、因果関係の有無のみではノイズ除去としては不十分である。本研究において、理想的な課題－効果文とは、課題が、特別な技術的特徴(STF: Special Technical Feature)を表している文である。ここで、特別な技術的特徴とは、特許の核となる「発明の先行技術に対する貢献を明示する技術的特徴」[特許庁 2016]を表す。

このような特別な技術的特徴を含む課題を抽出するために、2 つ目のノイズ除去では、特徴的な単語(キーワード)を課題に含まない課題－効果文候補はノイズとして除去する。なお、本研究において、特徴的な単語とは、しきい値以上の TF-IDF を持つ単語である。ここで、IDF は、提案手法の入力である探索範囲の全特許文書より算出する。

3.4 文章重要度に基づくノイズ除去

最後のノイズ除去の方法は、文章重要度に基づくノイズ除去である。本ノイズ除去では、一般的な表現や具体性に欠ける表現等を候補から取り除くことを目的とする。

例えば、典型的な効果の手がかり表現である「できる。」を考える。多くの場合、「〇〇できる。」は正しく発明の効果を表すものであるが、「特徴づけることができる。」といった表現も抽出される。このような表現は特許文書において一般的に用いられる表現であるが、具体的な効果を表すものではない。また、課題においても、「このような材料を用いることにより、」といった照応表現を用いて具体的な技術的内容を含まない不適切な課題が抽出されることがある。

以上のような効果や課題の共通の特徴は、ともに一般的な語で表現が構成される点にある。そこで、最後のノイズ除去では、課題－効果文の候補のうち、一般的な語のみで構成されるような文を除去する。具体的には、課題－効果文候補に含まれる語の TF-IDF の合計値を文章重要度とし、文章重要度がしきい値以下の場合に、当該文はノイズであると判定する。

上記のように、課題－効果文の抽出ステップでは、課題－効果文の候補を抽出した後、さらに 3 種類のノイズ除去手法を適用することで、探索範囲の各特許から課題－効果文を抽出する。

3.5 類似課題－非類似効果文の検索

次に、実際に新しい用途先となる特許を探す、類似課題－非類似効果文探索ステップについて説明する。本ステップは、前ステップによって抽出された課題－効果文から、クエリ(探索基準の課題－効果)と比較して、課題が類似し、かつ効果が類似していない課題－効果文を検索する処理である。

ここで、検索の対象となる課題－効果文は、元の特許文書と比較すると非常に短い。そのため、課題や効果の類似性を単純な one-hot 表現を用いて求めようとすると、多くの場合で類似していないと判定されてしまい、正しく類似課題－非類似効果文を検索することができない。

そこで、本研究では、課題－効果文を構成する語の分散表現を求め、課題(および効果)を構成する語の分散表現の和(以下、表現ベクトルと呼ぶ)を比較することで課題(および効果)の類似／非類似を判定する。具体的には、本研究では、word2vec¹によって語の分散表現を求める。また、表現ベクトルの類似性判定は、コサイン類似度がしきい値を超えたか否かによって判定を行う。

なお、本研究では、コサイン類似度による比較を行う前に、高 TF-IDF のキーワードによるフィルタリングを行い、類似度を算出する課題－効果文を絞り込んでいる。

以上の方法によって、提案手法は、探索範囲の課題－効果文から、クエリと類似課題－非類似効果を持つ文を抽出し、当該文を含む特許を出力する。すなわち、類似している課題を持つ特許において、課題と対応している効果が類似していない特許を抽出する。

4. 実験

本研究では提案手法について、2 つの実験を行った。1 つ目の実験は、実際に新しい発明の用途が探索可能であるか否かを確認する実験(以下、新規用途探索実験と呼ぶ)である。2 つ目の実験は、過去、実際に他の用途へ展開した発明をクエリとして、提案手法によって、その展開した用途が検出できるか否かを確認する実験(以下、他用途展開再現実験と呼ぶ)である。

なお、今回の実験では、探索対象の特許集合や各種パラメータは同一のものをを用い、クエリ(探索基点の課題－効果)のみ別のものをを用いた。両実験で共通のデータやパラメータを表 1 に示す。表 1 において、初期手がかり表現とは、坂地らの手法(2010)で最初に与える手がかり表現である。また、助詞 SVM の教師データとは、坂地らの手法(2011)で因果関係の有無を判定する SVM を学習するために用いた文である。また、重要単語のしきい値とは、重要単語によるノイズ除去において、単語がキーワードであるか否かを判定するためのしきい値である。また、文章重要度のしきい値とは、文章重要度によるノイズ除去において、課題－効果文がノイズか否かを判定する合計 TF-IDF のしきい値である。なお、本研究では、形態素解析器は MeCab²を(辞書は NEologd³を使用)、構文解析器は CaboCha⁴を用いた。

以下、各実験の詳細について述べる。

4.1 新規用途探索実験

最初の実験では、ある特定の用途の素材に関する特許 A についての課題－効果文をクエリとし、当該課題－効果と類似課題－非類似効果を持つ特許を出力し、その出力にどの程度他の用途先が含まれるかを確認した。

¹ <https://code.google.com/archive/p/word2vec/>

² <http://taku910.github.io/mecab/>

³ <https://github.com/neologd/mecab-ipadic-neologd>

⁴ <http://taku910.github.io/cabocha/>

表 1: 実験条件

探索対象	2013 年～2015 年の すべての公開特許公報
初期手がかり表現	[坂地 2010]と同じ
助詞 SVM の 教師データ	上記探索対象から人手で 判定した 6,300 文
重要単語のしきい値	3
文章重要度のしきい値	10
(word2vec) 学習アルゴリズム	CBOW
(word2vec) 次元数	300
課題類義語しきい値	0.7

実験の結果、特許 A の課題－効果文から、306 件の特許が抽出された。実際に 306 件のすべての特許を読み込み、他の用途先となっているかを人手で判定した結果、20 個の用途先を確認できた。

4.2 他用途展開再現実験

次の実験では、過去、実際に起こった用途展開例を用いて、その用途展開を再現できるか否かを検証した。本実験では、分野 Y の製品へ展開した分野 X の特許について、特許の課題－効果文から分野 Y を発見することができるか否かを確認した。

実験では、分野 X の特許 4 件の課題－効果文から、200 件の特許を抽出した。それぞれの特許の技術分野を判定した結果、図 5 のような結果となった。図 5 のように、分野 Y の特許を 6 件 (5%) 確認した。このように、用途展開した実例を再現することに成功した。

5. おわりに

本研究では、積極的に知財を活用するために、既存発明の技術的特徴を用いて、新しい効果を生み出すような用途先(展開先)を発見するための手法を提案した。

提案手法は、はじめに、大量の特許文書から発明の効果とその効果を達成するために必要であった技術的要素を表す文(課題－効果文)を抽出する。その後、提案手法は、抽出した課題－効果文の中から、探索基点とする特許(クエリ)の課題と類似し、かつ効果が類似していない文を抽出する。このようにすることで、類似した技術的特徴が生み出す、新しい効果を探索することが可能となる。

さらに、実際の特許文書を用いて提案手法を検証した結果、新しい用途先が探索可能であることを確認し、さらに、過去の用途展開事例を再現できることも確認した。これより、本研究で提案する用途探索手法は、有用であると考えられる。

なお、著者らは、自動的に新規用途先特許を探索する本提案手法を実装したシステムを開発している(図 6)。

また、提案手法は、多少の変更で、さらなる応用が可能である。例えば、提案手法の類似課題－非類似効果を探すステップを変更し、クエリと同一の技術分野の特許から類似した効果を持ち、かつ類似していない課題を持つ文(非類似課題－類似効果文)を抽出することで、クエリと同じ目的を、別の手段で解決するような特許を発見できる。このようにすることで、開発者は、発明を改良するための参考資料を容易に発見することができる。

以上のように、提案手法は、知財業務の様々な場面で活用が可能な手法であるが、取り組むべき課題も存在する。今後の課題を以下に示す。

- より大規模かつ詳細な実験

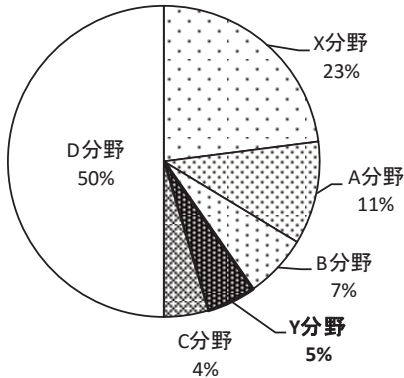


図 5: 他用途展開再現実験の実験結果



図 6: 用途先探索システム (CEES)

- 提案手法の再現率向上
今後、これらの課題に対する取り組みを行う予定である。

参考文献

[Nonaka 2010] Hirofumi Nonaka, Yusuke Suzuki, Hiroki Sakaji, Hiroyuki Sakai, Shigeru Masuyama: Extracting EFFECT Expressions Form Patent Documents, Proc. of the 25th International Technical Conference of Circuits/Systems, Computers and Communications (ITC-CSCC 2010), pp.729-731, 2010.

[石川 2007] 石川 大介, 石塚 英弘, 藤原 謙: 特許文献における因果関係を用いた類推による仮説の生成と検証 -ライフサイエンス分野を対象として-, 情報知識学会誌, Vol. 17, No. 3, pp. 164-181, 2007.

[坂地 2010] 坂地 泰紀, 野中 尋史, 酒井 浩之, 増山 繁: Cross-Bootstrapping : 特許文書からの課題・効果表現対の自動抽出手法, 電子情報通信学会論文誌. D, 情報・システム, Vol. 93, No. 6, pp. 742-755, 2010.

[坂地 2011] 坂地 泰紀, 増山 繁: 新聞記事からの因果関係を含む文の抽出手法, 電子情報通信学会論文誌. D, 情報・システム, Vol. 94, No. 8, pp. 1496-1506, 2011.

[特許庁 2016] 特許庁: 特許・実用新案審査基準 第 II 部 明細書及び特許請求の範囲 第 3 章 発明の単一性(特許法第 37 条), http://www.jpo.go.jp/shiryou/kijun/kijun2/pdf/tukujitu_kijun/02_0300.pdf, 2016 (2018 年 2 月 28 日アクセス).

[特許庁 2017] 特許庁: 知財人材スキル標準(version 2.0), https://www.jpo.go.jp/sesaku/kigyo_chizai/chizai_skill_ver_2_0.htm, 2017 (2018 年 2 月 28 日アクセス).

[日本特許 2011] 日本特許, 公開特許公報, 情報出力装置、情報蓄積装置、およびプログラム, 特開 2011-22668, 2011.